

数据挖掘-朴素贝叶斯

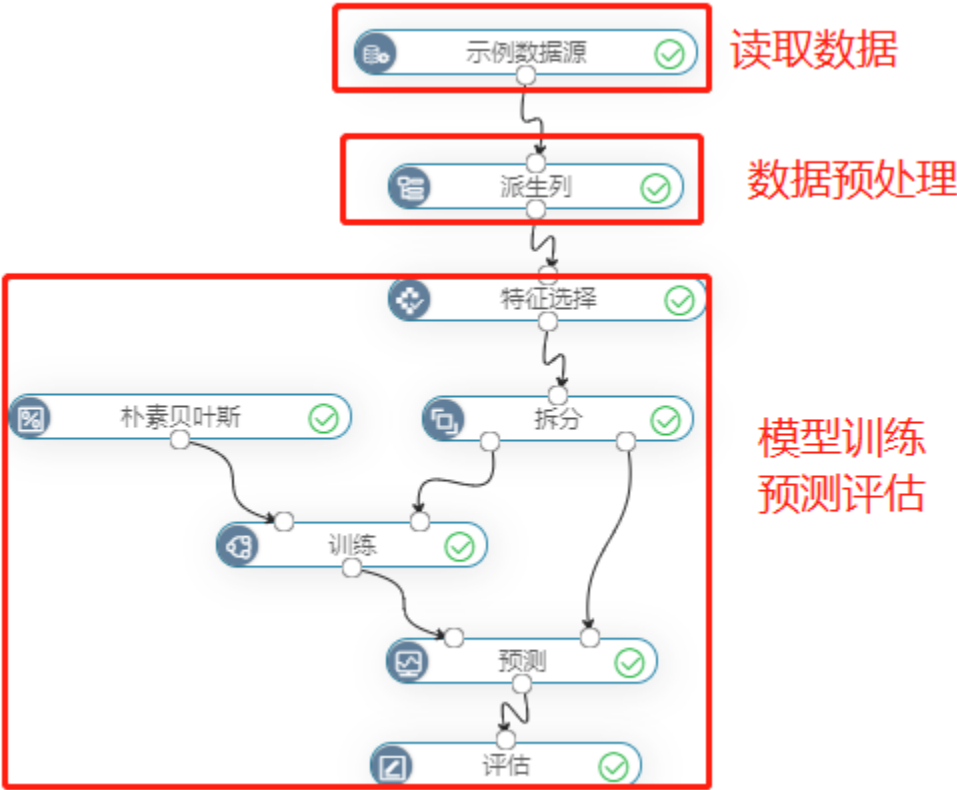
- 概述
- 示例
- 参数设置
 - 自动调参设置
 - 示例

概述

一种基于概率网络的分类算法，它在朴素贝叶斯定理的基础上取消了关于各属性关于类标号条件独立的苛刻条件，通过各类的先验概率计算待分类样本的后验概率，得到测试样本属于各类别的概率。它对于解决复杂设备不确定性和关联性引起的故障有很大的优势。

示例

使用“城市功能区识别”案例数据，预测城市功能区为专营商业区还是购物中心商业区两类。



其中派生列是为了衍生新的特征方便模型训练效果。

参数设置

朴素贝叶斯的参数如下：

参数名称	值	说明
自动调参设置	系统默认的各项参数值范围。	必须结合“启用自动调参”功能使用。系统将对设置指定或范围内的参数值循环调参，匹配出最优的组合。详情请参考 自动调参设置 。
启用自动调参	勾选该项，则系统自动调参数，不需要用户手工设置参数。	

模型类型	离散数据	根据特征数据的先验分布不同，可选择以下朴素贝叶斯算法： <ul style="list-style-type: none">• multinomial：多项式朴素贝叶斯，此选项为默认选项。• complement：补充朴素贝叶斯• gaussian：高斯朴素贝叶斯• bernoulli：伯努利朴素贝叶斯，对于小数量级的文本数据比较有优势。选bernoulli需要特征转化成0和1。
平滑参数	连续数据，必填	用于解决零概率问题，即测试样例的标签属性中，出现了模型训练过程中没有记录的值。
分类阈值	必填	多分类请用英文逗号隔开，且数量与分类数相同，例如：分三类，示例：8,9,10

自动调参设置

系统将对设置指定或范围内的参数值循环调参，匹配出最优的组合。

自动调参设置

拆分比例 *

0.09999999999999999

评估标准

precision

参数	指定值 *	范围 *	步数 *	是否使用指定值
模型类型	compler	complement		<input type="checkbox"/>
平滑参数	1	0.1 - 1	8	<input type="checkbox"/>

注意：勾选使用指定值时,不进行范围调参。

确定

取消

自动调参的方式分为两种：

- 指定值调参：指定一个固定的值进行自动调参。
- 范围调参：在指定的范围内进行自动调参。

设置项说明如下：

设置项		说明
拆分比例		将选择的数据拆分为两部分，一部分部分用于模型的评估，另一部分数据用于训练模型。
评估标准		用于选择数据的评估指标，包括：f1、precision、recall、accuracy、AUC(二分类)。 其中，评估标准“AUC(二分类)”仅对二分类问题生效。
参数		自动调参的参数项。
指定值调参	指定值	指定一个固定的值进行自动调参。 需要先勾选“是否使用指定值”才能配置。
	是否使用指定值	控制是否使用使用指定值进行调参。

范围调参	范围	<p>设置自动调参参数的范围。</p> <p>若运行速度比较慢，可将参数范围调小一点。</p> <p>注意：</p> <ul style="list-style-type: none"> 模型类型的范围支持多选。 平滑参数默认的范围提供了一个推荐值，并不是算法限制的最大值和最小值。
	步数	<p>进行范围调参时，在设置的范围内生成多少个参数值。</p> <p>示例：</p> <p>1) 范围为[3, 5]，步数为3时，生成的参数值：3, 4, 5</p> <p>2) 范围为[40, 100]，步数为4时，生成的参数值：40, 60, 80, 100</p>

示例

设置自动调参设置如图：

自动调参设置
 ×

拆分比例 *

评估标准

precision

参数	指定值 *	范围 *	步数 *	是否使用指定值
模型类型	<div>compler</div>	<div>complement</div>		<input type="checkbox"/>
平滑参数	<div>1</div>	<div>0.1</div> - <div>1</div>	<div>8</div>	<input type="checkbox"/>

注意：勾选使用指定值时,不进行范围调参。

确定

取消

在训练节点查看分析结果如图：

模型参数

超参数名称	超参数值
模型类型(选bernoulli需要特征转化成0和1)	complement
平滑参数	0.1